ITC19/ Performance Challenges for Efficient Next Generation Networks 1517-1526
LIANG X.J. and XIN Z.H.(Editors)
V.B. IVERSEN and KUO G.S.(Editors)
Beijing University of Posts and Telecommunications Press

User-level performance in WLAN hotspots

Franck Lebeugle [1] and Alexandre Proutiere [1]

France Telecom R&D, 38-40 rue du General Leclerc, 92794 Issy-les-Moulineaux, France
{franck.lebeugle,alexandre.proutiere}@francetelecom.com

Abstract. This paper presents a novel performance analysis of the downlink of WLANs carrying TCP traffic. We first provide a packet level analysis that accounts not only for the interaction between TCP and the IEEE 802.11 MAC protocol, but also for the different radio conditions experienced by various users. We then study the quality of service as perceived by users, i.e., we characterize the mean transfer time of data flows (web pages, files, etc.). The latter performance metric and its evaluation can then be used to design efficient dimensioning rules. The derived analytical results are validated through simulations.

1 Introduction

IEEE 802.11 [1] is the most popular standard for wireless LANs. To correctly dimension such networks, we need to understand the three-way relationship between radio capacity, traffic demand and realized quality of service. In this paper we consider this relationship under the assumption that traffic consists wholly of data transfers (Web pages, files,...), corresponding to the most common use of WLANs today. We seek to evaluate how the mean transfer time or, equivalently, the expected flow throughput depends on traffic intensity. This would allow a network designer to decide how many access points would be necessary at a hotspot, for example.

We consider a WLAN running the Distributed Coordination Function (DCF) MAC protocol and assume traffic consists of TCP flows transmitting data from the access point to users together with acknowledgement packets flowing in the reverse direction. The number of concurrent flows is a random process resulting from a potentially large user population initiating finite transfers according to a certain stochastic process. The proposed method can be applied to other traffic configurations, however, including flows transmitting data from users to the access point.

A first step in deriving flow-level performance results is to understand performance realized by DCF at packet level (i.e., when the number of data flows is fixed). This issue has been extensively addressed in the literature starting with the seminal paper by Bianchi [4]. Assuming a fixed number of saturated sources (i.e., the sources always have data to send) compete for the use of the channel, a decoupling assumption is applied to evaluate the stationary probability a source attempts to use the channel by solving a fixed-point equation.

The approach has been generalized to account for non-saturated sources and service differentiation realized with the IEEE 802.11e standard [6,14]. However, these results can not be directly applied to predict the throughput of TCP flows: unlike resources in wireline or traditional cellular access networks, the same WLAN medium is used for the transmission of both TCP packets and corresponding acknowledgements (acks). These transmissions interfere and the impact of ack feedback is not negligible.

This issue has already been studied to some extent through simulation [9,5]. A very simple analytical model is proposed in [12] to account for the interactions between TCP and DCF in case where all users experience the same radio conditions (i.e., they all have the same transmission rate, equal to 1, 2, 5.5 or 11 Mbit/s in case of IEEE 802.11b). Our model accounts more accurately for these interactions and is valid when users experience different radio conditions.

WLAN performance at flow level has so far received little attention. The authors of [11] give an explicit formula for the mean flow transfer time, but they do not account for the feedback traffic. In [12], the transmission of TCP acks is taken into account but the fact that all users have the same radio conditions greatly simplifies the analysis since performance is then insensitive to detailed traffic characteristics such as the distribution of flow sizes. In the more realistic setting where users have different radio conditions, the system looses this insensitivity property and an exact performance analysis appears intractable. To get around this difficulty, we apply stochastic comparisons (as in [3]) to derive an insensitive lower bound of the relevant performance parameter. This bound can be used to dimension realistic WLAN hotspots supporting data traffic generated by users with different radio conditions.

The considered MAC protocol and the system and traffic parameters are described in the next section. We then provide analytical formulas for the instantaneous throughput of TCP connections and apply these results to derive a conservative estimate of the mean flow transfer time. We conclude the paper by presenting numerical experiments that compare the analytical evaluations with the results of simulations.

2 Protocol and models

2.1 CSMA/CA and the DCF function

The fundamental access method of the 802.11 MAC is the Distributed Coordination Function known as Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) [1].

CSMA/CA Principles. A source running the DCF is either in a contention phase or in a transmission phase.
Contention phase: The source runs a back-off timer. It is set when the source becomes active, after a successful or unsuccessful packet transmission. At each of these epochs, the back-off timer is $SLOT \times U$, where U is an integer uniformly chosen in the range $[0, CW-1]$ and where CW is an integer called the contention window. CW lies between $CW_{\min}$ and $CW_{\max} = \alpha^m CW_{\min}$, where α refers to the multiplicative factor (equal to 2 in the IEEE 802.11b). For the first transmission or after a successful transmission, $CW = CW_{min}$. After an unsuccessful transmission, CW is multiplied by α if $CW \neq CW_{\max}$ or remains unchanged otherwise. When the medium is sensed idle during a slot, the source decrements its back-off timer, when it is sensed busy, the timer is frozen. When the channel is sensed idle again for a $DIFS$, the timer is reactivated. When it reaches zero, the source attempts to use the channel.
Transmission phase. When the back-off timer reaches zero the source sends a packet (including the PLCP preamble and all headers). If the receiver successfully receives the packet, it waits for a $SIFS$ and then transmits an acknowledgment to the source. When a packet is not acknowledged, the source considers that a collision occurred.

We have implicitly assumed that the duration of a collision involving a data packet is always greater than that of a collision involving TCP acks only (whatever the flow classes). This assumption is valid with the IEEE 802.11b parameters.

With the RTS/CTS mechanism, the duration of any collision is:

$$T_i^{col,rts} = DIFS + t^{pr} + \frac{\sigma_{RTS}}{R_i}.$$

2.3 Traffic assumptions and performance metrics

We consider I classes of TCP flows generated on the downlink from the access point to receivers. Flows of class i are generated according to a Poisson process of intensity ρ_i and are of sizes i.i.d. with general distribution and of unit mean. Packets corresponding to a class-i flow are transmitted at rate R_i. The number of active class-i flows is randomly varying and is denoted by x_i. In the following we use the notation $\boldsymbol{x} = (x_1, \ldots, x_I)$ and $x = \sum_i x_i$.

Remark: This approach can be generalized to the case where flows are generated within sessions, a session being an arbitrary succession of data flows separated by periods of inactivity, provided sessions are generated according to a Poisson process [2]. The approach can also be generalized to the case of a finite and fixed number of users alternating between flow transfers and periods of inactivity.

In case of data traffic, the performance perceived by users is defined by the mean time to transfer a flow, $E[S_i]$. Alternatively, we can measure performance through the mean flow throughput defined for class-i flows by $\gamma_i = 1/E[S_i]$.

3 Packet-level performance

We first provide a performance analysis at packet level in a general setting where x sources compete for the use of the radio channel. The analysis is similar to that used by Bianchi [4]. Then, we apply this analysis to compute the instantaneous throughput obtained by each TCP flow on the downlink.

3.1 The decoupling assumption and the fixed point approach

We consider x competing sources and compute the stationary probability a source attempts to use the channel at time epochs when the back-off timers are decremented.

The decoupling assumption. It proves extremely difficult to analyze the performance of CSMA/CA without any simplifying assumption. The basic assumption is that of decoupling: at the stationary regime, a given station will attempt the channel independently of the behavior of the other stations. Let p be the stationary probability a station attempts to use the channel. Then a station gains access at rate $r = p(1-p)^{x-1}$. This assumption, initially used by Bianchi [4], has been widely used by other authors. Its validity has never been addressed theoretically, but it surprisingly leads to tight approximations of the performance [4, 11]. The analysis focuses on the behavior of one particular station at the instants when the back-off timer of the station changes.

The fixed point approach. Under the decoupling assumption, the performance analysis is equivalent to solving a fixed point equation. In the stationary regime, each station attempts to

Note that the evolution of the process $y(t)$ depends on the numbers of flows of class i, x_i, through their sum x only. Now fix the state $y(t) = j$. When a transmission occurs with success (a data packet or a TCP ack), the probability that it corresponds to a class-i connection is x_i/x. When a collision occurs involving n sources, the duration of the collision is T_i^{col} with probability proportional to $(x_i/x)\binom{j}{n-1}$. The duration of the collision is $T_i^{col,ack}$ with probability proportional to

$$\alpha_n(i, \boldsymbol{x}) = \sum_{k=1}^{n \wedge x_i} \binom{x_i}{k} \binom{x_{(i+1)\rightarrow}}{n-k}, \tag{3}$$

where $x_{l\rightarrow} = \sum_{k=l}^{I} x_k$. Note that $\alpha_n(i, \boldsymbol{x})$ does not depend on $y(t)$ but on n only. We denote by $\beta_n(i, \boldsymbol{x}) = \alpha_n(i, \boldsymbol{x}) / \sum_j \alpha_n(j, \boldsymbol{x})$. The global stationary throughput of the system is given by:

$$\phi(\boldsymbol{x}) = \frac{\sum_{j=0}^{x-1} \pi_j^x \times G_j}{\sum_{j=0}^{x} \pi_j^x \times P_j}, \tag{4}$$

with:

$$G_j = p_{j+1}(1 - p_{j+1})^j \sigma, \tag{5}$$

for $j < x$:

$$\begin{aligned} P_j = {} & p_{j+1}(1 - p_{j+1})^j \sum_{i=1}^{I} \frac{x_i}{x}(T_i + jT_i^{ack}) + (1 - p_{j+1})^{j+1} SLOT \\ & + \sum_{n=2}^{j+1} p_{j+1}^n (1 - p_{j+1})^{j+1-n} \left(\binom{j}{n-1} \sum_i (x_i T_i^{col})/x + \binom{j}{n} \sum_i \beta_n(i, \boldsymbol{x}) T_i^{col,ack} \right)), \end{aligned} \tag{6}$$

and

$$P_x = p_x(1-p_x)^{x-1} \sum_i x_i T_i^{ack}/x + (1-p_x)^x SLOT + \sum_{n=2}^{x} p_x^n (1-p_x)^{x-n} \binom{x}{n} \sum_i \beta_n(i, \boldsymbol{x}) T_i^{col,ack}, \tag{7}$$

Finally, each TCP flow receives a throughput equal to $\phi(\boldsymbol{x})/x$. The fact that the instantaneous flow throughput does not depend on the the flow class has been pointed out in [9]. Actually, this does not imply that users with different radio conditions will experience the same QoS at flow level, as we will show in Sections 4-5.

Homogeneous radio conditions. When all TCP flows have the same data rate, the throughput of the system is given by (dropping index i):

$$\phi(x) = \frac{\sum_{j=0}^{x-1} \pi_j^x \times G_j}{\sum_{j=0}^{x} \pi_j^x \times P_j}, \tag{8}$$

with for $j < x$:

$$\begin{aligned} P_j = {} & p_{j+1}(1 - p_{j+1})^j (T + jT^{ack}) + (1 - p_{j+1})^{j+1} SLOT \\ & + \sum_{n=2}^{j+1} p_{j+1}^n (1 - p_{j+1})^{j+1-n} \left(\binom{j}{n-1} T^{col} + \binom{j}{n} T^{col,ack} \right)), \end{aligned} \tag{9}$$

Let $\phi_i(\boldsymbol{x}) = x_i\phi(\boldsymbol{x})/x$ be the service capacity of TCP flows of class i when the system is in state $\boldsymbol{x}$. In the following we denote by e_i the unit vector with 1 in component i and 0 elsewhere, for $i = 1, \ldots, I$. For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^I$, we use the notation $\boldsymbol{x} \leq \boldsymbol{y}$ if $x_i \leq y_i$, for all i. We denote by $\boldsymbol{X}(t)$ the process representing the evolution of the number of TCP flows of each class, when the system starts empty at time 0.

Monotonicity property. It is straightforward to verify that the following monotonicity property holds.

$$\frac{\phi_i(\boldsymbol{x})}{x_i} \geq \frac{\phi_i(\boldsymbol{y})}{y_i}, \quad \forall \boldsymbol{x} \leq \boldsymbol{y} : x_i > 0, \forall i. \tag{14}$$

The above property basically states that when there are fewer flows, each of them receives a greater instantaneous throughput.

Define the function $\check{\Phi}(\boldsymbol{x})$ as follows: $\check{\Phi}(0) = 1$ and for $\boldsymbol{x} > 0$,

$$\check{\Phi}(\boldsymbol{x}) = \frac{x!}{x_1! \ldots x_I!} \left(\prod_{n=1}^{x_I} \phi(ne_I) \times \prod_{n=1}^{x_{I-1}} \phi(ne_{I-1} + x_I e_I) \times \ldots \times \prod_{n=1}^{x_1} \phi(ne_1 + x_2 e_2 + \ldots + x_I e_I) \right)^{-1}. \tag{15}$$

Also define the capacities $\check{\phi}_i(\boldsymbol{x}) = \check{\Phi}(\boldsymbol{x} - e_i)/\check{\Phi}(\boldsymbol{x})$, for all i. Finally denote by $\check{\boldsymbol{X}}(t)$ the vector of the number of flows of each class at time $t \geq 0$ starting with an empty system at time 0, and with service capacities equal to $\check{\phi}_1, \ldots, \check{\phi}_I$. Then we have by construction [3]:

$$\forall i, \boldsymbol{x}, \phi_i(\boldsymbol{x}) \geq \check{\phi}_i(\boldsymbol{x}) \quad \text{and} \quad \forall t \geq 0, \boldsymbol{X}(t) \leq \check{\boldsymbol{X}}(t). \tag{16}$$

The stationary distribution $\check{\pi}$ of the process $\check{\boldsymbol{X}}(t)$ and the corresponding mean flow throughputs $\check{\gamma}_i$ are known, because $\check{\boldsymbol{X}}(t)$ has been constructed so as to be a Whittle process [14]. This stationary distribution and the flow throughputs are insensitive to flow size distributions and are given by:

$$\check{\pi}(\boldsymbol{x}) = \check{\pi}(0)\check{\Phi}(\boldsymbol{x}) \prod_{i=1}^{I} \rho_i^{x_i}, \quad \check{\gamma}_i = \frac{\rho_i}{E[\check{X}_i]}, \forall i. \tag{17}$$

In view of (16): $\forall i, \quad \check{\gamma}_i \leq \gamma_i$. We have thus derived an insensitive lower bound of the flow-level performance, useful for dimensioning purposes.

5 Numerical experiments

We now apply the results derived above to evaluate the performance in WLANs under the IEEE 802.11b standard. The values of the system parameters are provided in [1].

Packet-level performance. We first fix the number of TCP flows. The left side of Figure 1 shows the global instantaneous downlink throughput $\phi(x)$ given by (8) as a function of the number of TCP flows and when users experience the same radio conditions (the data rate is fixed at 11 Mbit/s). As mentioned earlier this throughput is well approximated by ϕ^* even for small number of flows. The right side of Figure 1 shows the throughput $\phi(\boldsymbol{x})$ given by (4) with 10 TCP flows

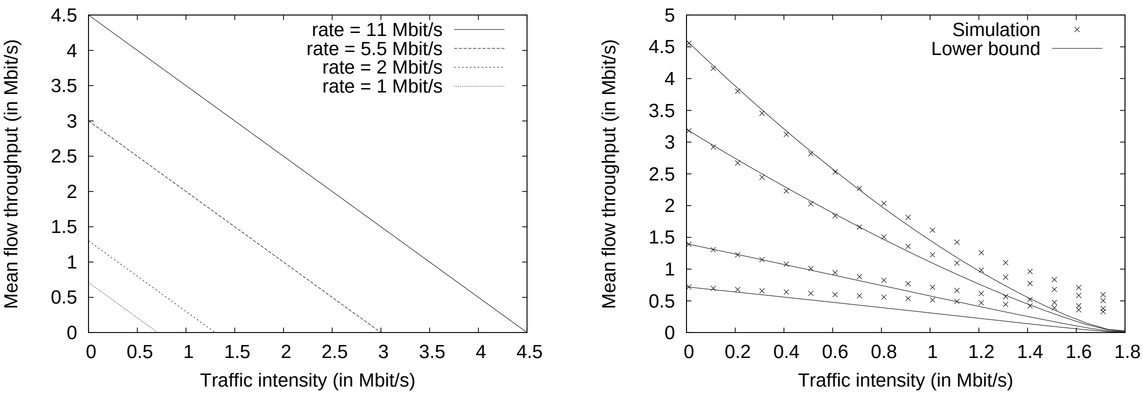

Fig. 2. Flow-level performance in case of homogeneous (left) and heterogeneous (right) radio conditions.

References

1. IEEE Standard for Wireless LAN Medium Access Protocol and Physical Layer Specifications, IEEE std., August 1999.
2. S. Ben Fredj, T. Bonald, A. Proutiere, G. Regnie, J. Roberts, Statistical bandwidth sharing: a study of congestion at flow level, in: *Proc. of ACM Sigcomm*, 2001.
3. T. Bonald, A. Proutiere, On stochastic bounds for monotonic processor sharing networks, Queueing Systems 47 (2004) 81-106.
4. G. Bianchi, Performance Analysis of the IEEE 802.11 Distributed Coordination Function, IEEE Journal on Selected Areas in Communications 18-3 (2000) pp 535-547.
5. D. Chapon, F. Lebeugle, Modelling of the IEEE 802.11 MAC layer via ns simulations, *France Telecom Research Report*, 2004. (Soon available for download).
6. K. Chen, Y. Xue, K. Nahrstedt, On setting TCP's congestion window limit in mobile ad hoc networks, in: *Proc. of IEEE ICC, Anchorage*, 2003.
7. I. Chlamtac, G. Zeng, A Finite Queueing Model for IEEE 802.11 MAC Protocol, in: *Proc. of IEEE VTC*, Italy, May 2004.
8. P. Gupta, Y. Sankarasubramaniam, A. Stolyar, Random-access Scheduling with Service Differentiation in Wireless Networks, in: *Proc. of IEEE Infocom*, 2005.
9. M. Heusse, F. Rousseau, G. Berger-Sabbatel, A. Duda, Performance Anomaly of 802.11b, in: *Proc. of IEEE Infocom*, 2003.
10. R. Jiang, V. Gupta, C.V. Ravishankar, Interactions between TCP and the IEEE 802.11 MAC protocol, in: *Proc. of IEEE DISCEX, Washington D.C., USA*, April 2003.
11. A. Kumar, E. Altman, D. Miorandi and M. Goyal, New Insights from a Fixed Point Analysis of Single Cell IEEE 802.11 WLANs, in *Proc. of IEEE Infocom*, 2005.
12. R. Litjens, F. Roijers, H. Van Den Berg, R.J. Boucherie, M. Fleuren, Analysis of flow transfer times in IEEE 802.11 wireless LANs, Annals of Telecommunications 59 (2004) pp 1407-1432.
13. D. Miorandi , A. A. Kherani, E. Altman, A queueing model for HTTP traffic over IEEE 802.11 WLANs, in: *Proc. of ITC Specialist Seminar on Performance Evaluation of Wireless and Mobile Systems*, University of Antwerp, Belgium, 2004.
14. R. Serfozo, *Introduction to Stochastic Networks*, Springer-Verlag, 1999.
15. H. Zhu, I. Chlamtac, An Analytical Model for IEEE 802.11e EDCF Differential Services, in: *Proc. of IEEE ICCCN, Dallas*, October 2003.